基于异常数据预处理和自适应估计的 WSN 数据融合算法 *

郑宝周，吴莉莉，李富强，袁　超

(河南农业大学 理学院，郑州 450002)

摘　要：针对无线传感器网络(wireless sensor network，WSN)存在节点能量受限、测量精度低、生存期短等问题，提出一种基于异常数据预处理和自适应估计加权融合算法(abnormal data-preprocessing adaptive estimation weighting fusion，ADAEWF)。为了提高算法可靠性，提出了基于异常数据检测、简单多数原则和节点综合支持度函数的数据预处理机制；为了减小测量误差对融合精度的影响，基于分批估计和自适应理论对节点测量值进行自适应估计加权数据融合；然后，建立了 WSN 仿真模型，并分别获得了 ADAEWF、自适应预测加权数据融合算法(adaptive forecast weighting data fusion，AFWDF)和算术平均值法下融合结果的均方误差和网络有效生存期。仿真结果显示：ADAEWF 算法融合精度和网络有效生存期均优于 AFWDF 和算术平均值法，表明 ADAEWF 算法在提高融合数据有效性、网络有效生存期和融合精度方面具有优越性。

关键词：无线传感器网络；异常数据预处理；自适应分批估计；数据融合

中图分类号：TP393.03　　doi: 10.3969/j.issn.1001-3695.2018.07.0345

Data fusion algorithm based on abnormal data-preprocessing and adaptive estimation in WSN

Zheng Baozhou, Wu Lili, Li Fuqiang, Yuan Chao

(*College of Sciences Henan Agricultural University*, *Zhengzhou* 450002, *China*)

Abstract: This paper aimed at the WSN's problems, such as limited node's energy, low precision of measurement and short effective lifetime of network, etc, proposed a data fusion algorithm based on abnormal data preprocessing and adaptive estimation, ADAEWF. In order to improve the algorithm's reliability, this paper firstly proposed the data preprocessing mechanism based on abnormal data detection, simple majority principle and node's comprehensive support function. To reduce the measurement error's effect on the fusion precision, this paper proposed an adaptive estimation weighted data fusion algorithm. Then, this paper established the WSN's simulation model and obtained the mean square error and network's effective lifetime under ADAEWF, AFWDF and arithmetic mean value algorithm, respectively. Simulation shows that, In terms of both fusion accuracy and network's effective lifetime, ADAEWF outperforms AFWDF and arithmetic mean value algorithm, and ADAEWF algorithm has superiority in improving the effectiveness of data fusion, network lifetime and fusion accuracy.

Key words: wireless sensor network; abnormal data preprocess; adaptive batch estimation; data fusion

0　引言

无线传感器网络(WSN)由大量具有感知、计算和无线通信能力的传感器节点组成，具有节点体积小、成本低、多跳自组网和感知区域大等特点。WSN 能够有效地对环境信息进行监测预警，已广泛应用于智能农业、环境监测和国防军事等领域。WSN 属于资源受限网络，其性能受节点的能量、计算能力和存储空间等因素制约[1]。通常在感知区域部署大量同构传感器节点，对信息进行周期性采集和传输，WSN 在监测信息的同时产生了大量重复性冗余数据。一方面，处理和传输冗余数据浪费了有限的能量和网络带宽；另一方面，干扰因素和传感器测量精度导致系统监测结果存在误差，随机出现的故障节点也在一定程度上降低了系统的可靠性。如何减少冗余数据、降低节点能耗、提高网络可靠性并延长其有效生存期，成为 WSN 研究中非常重要的问题[2]。

多传感器数据融合技术对具有一定冗余度的数据进行融合，能够有效减少数据传输、降低节点能耗和提高监测结果的准确性[3]。文献[1]用信息熵反映传感器数据分布的统计特性，最大

收稿日期：2018-07-13；**修回日期：**2018-08-27　**基金项目：**国家自然科学基金资助项目（61703146）；河南省科技攻关项目（172102210043）；河南省高等学校重点科研资助项目（15A510028）；河南农业大学科技创新基金项目（KJCX2015A17，KJCX2016A09）

作者简介：郑宝周（1979-），男，河南桐柏人，讲师，硕士，主要研究方向为无线传感器网络、信息检测及控制技术（baozhou120@163.com）；吴莉莉（1977-），女，副教授，博士，主要研究方向为传感技术、信号处理和模式识别；李富强（1982-），男，副教授，博士，主要研究方向为事件触发网络控制系统、多智能体系统；袁超（1961-），男，副教授，主要研究方向为传感器技术.

寻优确定数据融合上下限值，分别对节点数据和簇内数据进行两次融合。该算法融合度受限于信息熵值，节能有限。文献[2]提出的 AFWDF 算法基于节点数据在时间上的相关性建立预测模型，簇首根据特征值和预测值计算数据的可信度和权重，对簇内节点数据进行加权融合。AFWDF 需要在簇首和节点上建立预测模型，运算量大，融合精度和节点能量使用效率仍有提升空间。文献[4]基于模糊证据理论和模糊数学原理提出了新的节点间支持度计算方法，借鉴信任分配思想将测量值转换为对应的证据，通过证据组合规则得到融合结果。现有 WSN 数据融合算法共同点是：节点以固定周期采集、发送数据并启动融合算法[5~7]。实际应用中，监测对象大多是缓慢变化的物理量，如环境中的温度、湿度等信息，若无特殊事件发生，节点采集的数据大部分属于重复数据。然而，系统关注的主要是参数的变化情况，即在感知区域内是否有重要事件发生[8~10]。如果 WSN 长期处于高能耗的兴奋状态监测冗余数据，实际意义并不大。

异常数据在 WSN 监测系统中具有重要应用意义。异常数据是指明显偏离传感器正常工作模式的测量数据。异常数据的产生，意味着感知区域内发生重要监测事件，或者，外部干扰、节点故障等非正常因素导致传感器数据出现异常。无论是重要事件的发生还是故障节点的出现，均应受到足够的重视。外部事件引起的异常数据，系统应及时准确地采集、融合并上报监测结果；外界干扰和节点故障造成的异常数据，系统应进行识别、检测并做相应处理，避免对融合结果准确度造成影响[11]。

本文基于数据驱动思想，提出一种基于异常数据预处理的自适应估计加权融合算法(ADAEWF)。仿真结果表明，ADAEWF 算法能够有效降低 WSN 冗余数据量，延长网络有效生存期，并具有较高的融合精度。

1　网络模型及算法思想

如图 1 所示，WSN 由监测中心 PC 机、网关和多个无线传感器节点组成。监测系统以一阶无线电模型为 WSN 的能耗模型，采用能耗均衡的自适应分簇算法 (energy balanced adaptive clustering algorithm，EBACA)将网络分成若干个簇，每个簇包含簇首和若干个传感器节点[12]。传感器节点负责数据的采集、检测、传输以及节点状态和外部事件的监测；簇首节点收集簇内传感器节点数据、运行 ADAEWF 融合算法，并将融合结果上传至网关；网关将各簇数据融合结果转发至监测中心 PC 机。如果用户发出数据查询命令，则网关、簇首依次接收并转发查询命令至各传感器节点。

ADAEWF 算法思想如下：在自动监测模式下，以有效异常数据作为融合算法驱动源；传感器节点周期性采集数据，通过滑动窗口局部事件监测 (Sliding Window Local Event Detection, SW-LED)算法对异常数据进行检测[12]，将持续性异常数据发送至簇首；簇首根据简单多数原则和节点数据的综合支持度判断异常数据的有效性；如果异常数据数量达到刺激强度阈值，簇首触发自适应估计加权数据融合算法，最后簇首将融合结果上传至网关和监测中心 PC 机。在用户查询模式下，ADAEWF 算法将用户的查询请求视为重要监测事件，用户发出“数据查询”命令，则网关、簇首向网络广播“强制发送数据”命令，簇内各传感器节点采集当前数据并发送至簇首，簇首无条件触发自适应估计加权数据融合算法，最后将数据融合结果上传至网关和监测中心 PC 机。ADAEWF 算法流程如图 2 所示。

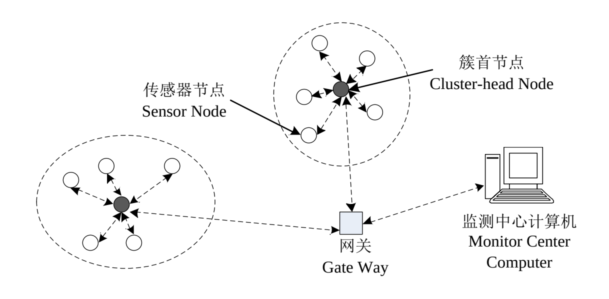


图 1　无线传感器网络结构图

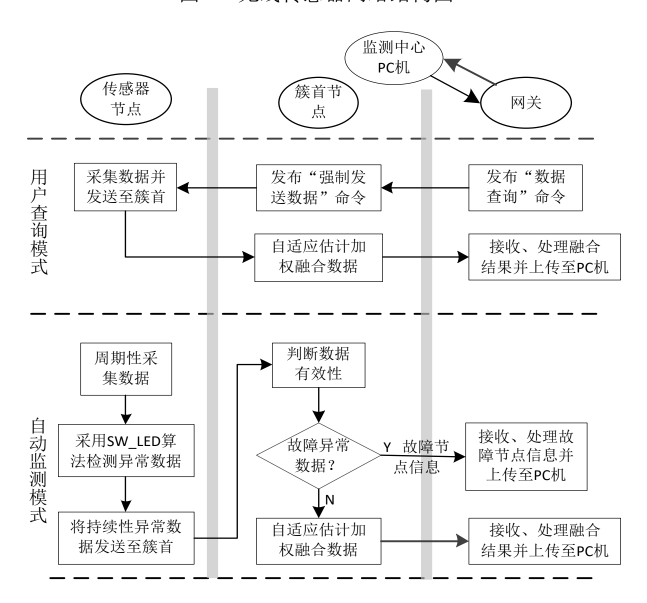


图 2　ADAEWF 算法流程图

2　ADAEWF 算法

2.1　异常数据预处理

传感器节点产生的异常数据分为三种：a)突发的外部干扰或短暂性故障引起的瞬时性异常数据；b)节点剩余能量不足、硬件故障或者软件缺陷引起的故障性异常数据；c)外部事件使环境信息发生较大幅度变化，导致多个节点采集的真实数据发生“异常”，称为有效异常数据[14]。下面介绍不同异常数据的处理方法。

2.1.1 瞬时性异常数据预处理

瞬时性异常数据出现频率较高，但不能准确反映真实环境信息，如果参与数据融合，会降低融合结果的准确度。系统采用 SW-LED 算法排除瞬时性异常数据参与融合运算[15]。

定义 1　对于节点数据 $x(t)$，若 $|x(t)-d|\ge\varepsilon$，则 $x(t)$ 为异常数据。其中，d 为该节点上一次向簇首发送的异常数据，ε 为异常数据判定阈值，其值取决于监测系统实际应用需求。

根据定义 1，传感器节点对 $x(t)$ 做二值化判断：若 $x(t)$ 为异常数据，则记 $x(t)$ 转换结果为 1，否则转换结果为 0；如果 $x(t)$ 转换结果为 1，进一步统计前面 $W-1$ 个数据转换结果之和是否达到 $0.5W$，达到 $0.5W$ 意味着滑动窗口内半数以上测量值为异常数据，取最终转换结果 $y(t)=1$，否则 $y(t)=0$，转换公式如式(1)所示。

$$y(t)=\begin{cases}1 & \sum_{j=t-W+1}^{t} x(j)\geq 0.5W \\ 0 & \sum_{j=t-W+1}^{t} x(j)<0.5W\end{cases} \tag{1}$$

其中：W 表示滑动窗口大小，如果 $y(t)=0$，意味着窗口内异常数据的数量达不到半数，$x(t)$ 被视为瞬时性异常数据，无需进一步处理；如果 $y(t)=1$，意味着窗口内半数以上测量值为异常数据，视 $x(t)$ 为持续性异常数据，节点将 $x(t)$ 传送至簇首。

2.1.2 故障性异常数据和有效异常数据预处理

随着系统的运行，节点剩余能量不断减少，同时节点软件、硬件也会受到各种不确定因素的影响，WSN 内会随机出现故障节点。故障节点在每个采集周期内持续的产生异常数据，称为故障性异常数据。一方面，故障性异常数据虽然能够通过 SW-LED 算法的检测，但不能反映真实的环境信息，如果参与融合算法，会降低融合结果的准确性；另一方面，持续存在的故障性异常数据会频繁地触发融合算法，其他节点则被动参与数据融合，浪费了节点能量和网络带宽。因此，簇首应对故障性异常数据进行判断识别，并将故障节点 ID 及状态信息发送至网关和和监测中心 PC 机。故障节点往往是随机出现的，位置相邻的多个节点同时出现故障的概率很小。然而，如果感知区域内发生重要监测事件，邻近的多个节点测量值会出现较大变化，比如：刮风下雨会使农田温度、湿度大幅度变化，导致传感器测量值出现异常。监测事件引起的异常数据真实反映了环境信息的变化，称为有效异常数据。

综上，对于故障性异常数据，簇首应将节点 ID 及状态信息上报至网关和监测中心 PC 机；对于有效异常数据，簇首应启动融合算法，对所有节点数据进行融合处理。

定义 2　簇首刺激强度阈值 γ，表示簇首兴奋门限。设簇内节点数为 m，根据简单多数原则将 γ 设为 $0.4m$。

设簇首收到 l 个节点的异常数据，如果 $l\geq\gamma$，簇首受到异常数据的刺激较强，认为感知区域内发生重要监测事件，簇首进入兴奋状态并触发数据融合算法；如果 $l<\gamma$，认为仅有少量节点产生故障性异常数据，簇首将故障节点信息发送至网关。

2.1.3 数据有效性判断

引入群体支持度方法对数据有效性进行判断：为每个节点数据引入一个支持度，支持度的大小体现了邻近节点对该数据有效性的支持程度。设多个节点测量同一个参数，节点 i 的测量值 a_i 有效性越高，则 a_i 被其他节点支持的程度就越高，a_i 为有效数据的可能程度也就越高。本文用支持度函数 $\sup(a,b)$ 表示数据 b 对 a 的支持程度。

Yager 提出支持度函数需满足以下三个条件：

a) $\sup(a,b)\in[0,1]$；

b) $\sup(a,b)=\sup(b,a)$；

c)若 $|a-b|<|x-y|$，则 $\sup(a,b)>\sup(x,y)$。

Yager 还给出了高斯型支持度函数：

$$\sup(a,b)=K\times e^{[-\beta(a-b)^2]},K\in[0,1],\beta\geq 0$$

高斯型支持度函数能够逼真地反映数据间的支持度关系，但是指数运算会占用节点大量的硬件资源，并不适用于资源受限的 WSN。本文采用基于灰色接近关联度的新型支持度函数来描述数据间的支持度[16]。

定义 3　数据 x_i 与 x_j 之间的支持度为

$$s_{ij}=D(x_i,x_j,K,\beta)=\frac{K}{1+\beta\times|x_i-x_j|\times|x_i-x_j|\times|x_i-x_j|\times|x_i-x_j|} \tag{2}$$

其中：$K\in[0,1]$，K 值决定了支持度函数的幅值；$\beta\geq 0$，其值越大，支持度函数衰减越快，称 β 为衰减因子；根据实际监测需求对 K 和 β 进行调节。式(2)满足 Yager 提出的三个必要条件，能够很好的逼近高斯型支持度函数曲线，因无需进行指数运算，节省了节点能量和硬件资源。

设 n 个节点测量同一参数，由式(2)计算支持度矩阵 S。

$$S=\begin{bmatrix}s_{11} & s_{12} & \cdots & s_{1n}\\ s_{21} & s_{22} & \cdots & s_{2n}\\ \vdots & \vdots & & \vdots\\ s_{n1} & s_{n2} & \cdots & s_{nn}\end{bmatrix} \tag{3}$$

定义综合支持度 s_i 表示数据 x_i 被其他节点综合支持的程度：

$$s_i=\sum_{j=1}^{n}s_{ij} \tag{4}$$

s_i 越大，则 x_i 与其他多数节点测量值越接近，x_i 的有效性越高；反之，s_i 偏离多数节点数据，其有效性越低。节点数据 x_i 有效性定义如下：

定义 4　数据有效性判定阈值 δ，若节点数据 x_i 综合支持度 $s_i\geq\delta$，则 x_i 为有效数据；否则，x_i 视为无效数据。

根据定义 4 对所有节点数据进行判断，有效数据参与数据融合，无效数据则无需参于数据融合，否则会降低融合结果的准确度。

2.2　分批估计融合算法

分批估计融合算法源于递推估计，通过对同一参数的多个测量值进行融合处理，能够有效降低测量误差对融合结果的影响，获取准确的测量结果[17]。

设 n 个节点对同一环境参数进行监测，根据相邻节点不同组的原则将节点分成 k 组，每组所含节点数可相同也可不同[17]。第 p 组节点采集的数据记为 $x_{p1},x_{p2},\cdots x_{pn_p}$，$p=1,2,\cdots k$，$n_p$ 为该组节点数量，节点数据平均值 $\bar{x}_p$ 及标准差 $\hat{\sigma}_p$ 为

$$\bar{x}_p=\frac{1}{n_p}\sum_{i=1}^{n_p}x_{pi},\quad p=1,2,\cdots k, \tag{5}$$

$$\hat{\sigma}_p = \sqrt{\frac{1}{n_p - 1}\sum_{i=1}^{n_p}(x_{pi} - \bar{x}_p)^2} \tag{6}$$

分批估计融合算法对节点当前测量值进行融合，无需考虑历史监测信息，认为历史数据的标准差 $\hat{\sigma}_- = \infty$，即 $(\hat{\sigma}_-)^{-1} = 0$。由分批估计理论可得融合结果如式(7)所示[17]。

$$\begin{aligned} x_+^{'} &= \left[\hat{\sigma}_+(\hat{\sigma}_-)^{-1}\right]\times x_- + \left[\hat{\sigma}_+ H^T R^{-1}\right]\times \bar{x} \\ &= \left[\hat{\sigma}_+ H^T R^{-1}\right]\times \bar{x} \end{aligned} \tag{7}$$

其中：$\bar{x}$ 为平均值矩阵，x_- 为上次数据融合结果，$\hat{\sigma}_+$ 为融合结果的方差，R 为测量噪声的协方阵，H 为测量方程的系数矩阵。各参数表达式如式(8)~(11)所示。

$$\hat{\sigma}_+ = \left[(\hat{\sigma}_-)^{-1} + H^T R^{-1} H\right]^{-1} \tag{8}$$

$$H = [1\ \ 1\ \ 1\cdots 1]^T \tag{9}$$

$$\bar{x} = [\bar{x}_1\ \ \bar{x}_2\ \ \cdots \bar{x}_k]^T \tag{10}$$

$$R = \begin{bmatrix} \hat{\sigma}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \hat{\sigma}_k^2 \end{bmatrix} \tag{11}$$

将式(9)~(11)和 $(\hat{\sigma}_-)^{-1} = 0$ 代入式(7)(8)，可得方差和分批估计融合结果如式(12)(13)所示。

$$\hat{\sigma}_+ = \left[(\hat{\sigma}_-)^{-1} + H^T R^{-1} H\right]^{-1} = \left[\sum_{p=1}^{k}\frac{1}{\hat{\sigma}_p^2}\right]^{-1} \tag{12}$$

$$x_+^{'} = \left[\sum_{i=1}^{k}\frac{1}{\hat{\sigma}_i^2}\right]^{-1}\sum_{p=1}^{k}\frac{\bar{x}_p}{\hat{\sigma}_p^2} = \sum_{p=1}^{k} w_p \bar{x}_p \tag{13}$$

2.3 自适应估计加权融合算法

为减小测量误差对融合精度的影响，将自适应理论与分批估计融合算法相结合，计算节点数据与分批估计融合结果的相对方差，根据相对方差调整节点数据的权值并进行自适应估计加权数据融合。图 3 为融合算法流程，算法步骤简述为：首先对节点分组，计算各组数据的均值 $\bar{x}_p$、标准差 $\hat{\sigma}_p$ 和融合权值 w_p，进一步求解分批估计融合结果 $x_+^{'}$；然后计算节点数据 x_{pi} 与 $x_+^{'}$ 的相对方差 $\bar{s}_p^2$，根据 $\bar{x}_p$、$\hat{\sigma}_p$ 和 $\bar{s}_p^2$ 计算每组数据的修正因子 a_p；最后，用修正后的权值 $w_p a_p$ 对数据进行二次加权融合，得到最终融合结果 x_+。

第 p 组节点数据融合权值 w_p、分批估计融合结果 $x_+^{'}$、相对方差 $\bar{s}_p^2$ 由式(14)-(16)求解。

$$w_p = \frac{1}{\hat{\sigma}_p^2}\left[\sum_{i=1}^{k}\frac{1}{\hat{\sigma}_i^2}\right]^{-1} \quad p = 1,2,\cdots k \tag{14}$$

$$x_+^{'} = \left[\sum_{i=1}^{k}\frac{1}{\hat{\sigma}_i^2}\right]^{-1}\times\sum_{p=1}^{k}\frac{\bar{x}_p}{\hat{\sigma}_p^2} = \sum_{p=1}^{k} w_p \bar{x}_p \tag{15}$$

$$\bar{s}_p^2 = \frac{1}{n_p}\sum_{i=1}^{n_p}(x_{pi} - x_+^{'})^2 \tag{16}$$

由多传感器加权数据融合算法可知，为消除测量误差，融合权值与估计方差应成反比，并且各权值相加和始终为 1，因此，a_p 与 w_p 满足式(17)~(19)三个约束条件[3]。

$$a_1 : a_2 : \cdots a_k = \frac{1}{\bar{s}_1^2} : \frac{1}{\bar{s}_2^2} : \cdots \frac{1}{\bar{s}_k^2} \tag{17}$$

$$\sum_{p=1}^{k} w_p = 1 \tag{18}$$

$$\sum_{p=1}^{k} w_p a_p = 1 \tag{19}$$

基于上述分析，修正因子 a_p 如式(20)所示。

$$a_p = \frac{1}{\bar{s}_p^2}\cdot\frac{1}{\sum_{i=1}^{k}\frac{w_i}{\bar{s}_i^2}} = \frac{1}{\bar{s}_p^2}\cdot\left[\left(\sum_{j=1}^{k}\frac{1}{\hat{\sigma}_j^2}\right)^{-1}\times\left(\sum_{i=1}^{k}\frac{1}{\bar{s}_i^2}\cdot\frac{1}{\hat{\sigma}_i^2}\right)\right]^{-1} \tag{20}$$

调整后的权值 $a_p w_p$ 代入式(15)，可求出自适应估计加权融合结果如式(21)所示。

$$\begin{aligned} x_+ &= \left[\sum_{i=1}^{k}\frac{1}{\hat{\sigma}_i^2}\cdot\frac{1}{\bar{s}_i^2}\right]^{-1}\cdot\sum_{p=1}^{k}\bar{x}_p\left(\frac{1}{\hat{\sigma}_p^2}\cdot\frac{1}{\bar{s}_p^2}\right) \\ &= \sum_{p=1}^{k} a_p w_p \bar{x}_p \end{aligned} \tag{21}$$

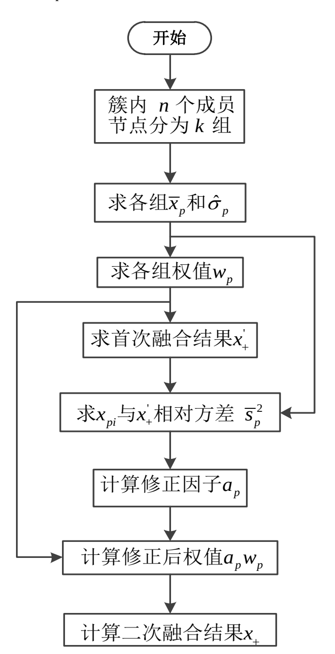


图 3 自适应估计加权融合算法流程图

2.4 ADAEWF 运行机制

角色不同的节点其功能行为亦不相同，为便于描述 ADAEWF 算法运行机制，定义如下三个状态参数：异常数据产生标志位 mTag、簇首兴奋标志位 hTag 和数据强制发送标志位 send[11]。

定义 5 异常数据产生标志位 mTag，反映簇成员节点是否产生非瞬时性异常数据，0 为否，1 为是。

定义 6 簇首兴奋标志位 hTag，反映簇首是否处于兴奋状态。hTag=0，表示簇首接收的异常数据个数小于 γ，簇首处于抑制状态；hTag=1，异常数据个数超过 γ，簇首处于兴奋状态；采集周期结束，hTag 重置为 0。

定义 7 数据强制发送标志位 send，反映节点是否收到"强

制发送数据”命令。send=1，表示节点收到“强制发送数据”命令，节点将测量值发送至簇首；send=0，表示节点未收到发送数据命令。

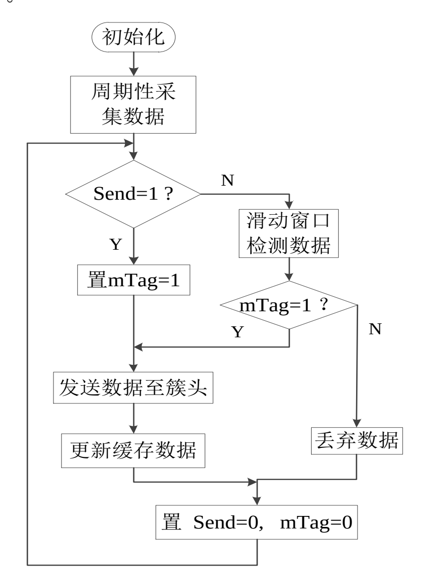


图 4　簇成员节点运行机制图

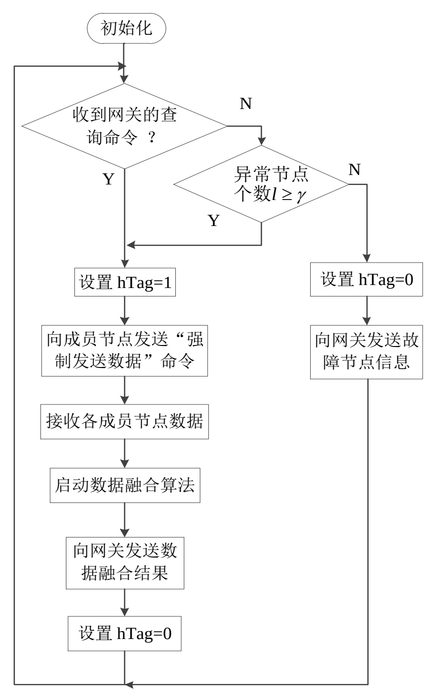


图 5 簇首节点运行机制图

簇成员节点动作行为如图 4 所示：①收到“强制发送数据”命令，节点处于兴奋状态，则置 send=1、mTag=1，节点将测量值 x_i^k 发送至簇首并更新缓存中数据；②如果 send=0，根据定义 1 判断如果 x_i^k 为瞬时性异常数据，则置 mTag=0 并将 x_i^k 丢弃不做处理；否则，节点置 mTag=1 并将 x_i^k 发送至簇首同时更新缓存中数据；③数据采集周期结束，重置 mTag、send 为初始值 0。

簇首动作行为如图 5 所示：a)如果收到网关转发的“数据查询”命令，簇首处于兴奋状态并置 hTag=1；b)根据定义 2 判断，如果异常数据个数 $l \geq \gamma$，簇首处于兴奋状态并置 hTag=1；c)如果 $l < \gamma$，簇首处于抑制状态置 hTag=0，并将故障节点信息发送至网关；d)如果 hTag=1，簇首广播“数据强制发送”命令，各成员节点发送数据，接收数据后簇首启动融合算法，并将融合结果发送至网关并将 hTag 恢复至初始值 0。

3　仿真实验结果及分析

为了分析和比较融合算法性能，本文基于 NS2 平台进行仿真实验，对 ADAEWF 算法、AFWDF 算法和算术平均值法的融合误差和网络有效生存期进行对比分析。实验背景设定为：在 200m×400m 农田内，均匀部署 200 个传感器节点对农田温度变化进行监测；节点行、列间距均为 10m，网关节点坐标为 (100,200)；节点故障、人工灌溉、天气变化等情况均会导致异常数据的产生；采用能耗均衡的 EBACA 算法对网络节点分簇，选用一阶无线电模型作为 WSN 的网络能耗模型[9]；其他仿真参数如表 1 所示。

表 1 仿真参数

参数 Parameter	参数值 Parameter Values
E_{elec} / $\mathrm{nJ \cdot bit^{-1}}$	50
$E_{init}(i)$ / J	2
d_0 / m	100
ε_{fs} / $\mathrm{pJ \cdot m^2 \cdot bit^{-1}}$	10
ε_{mp} / $\mathrm{pJ \cdot m^4 \cdot bit^{-1}}$	0.001
l / bit	2000
ε / ℃	1
γ	4

在函数 $T = 15 + 15\sin(2\pi ft)$ 上叠加方差为 5 的 0 均值白噪声信号，模拟农田温度变化以及电磁干扰、零点漂移等因素对节点传感器测量值的影响；随机选取一定比例节点，将节点数据设为 0 来模拟网络中随机出现的故障节点；调整 f 可改变异常数据出现的频率，f 越大，异常数据出现频率越高。

3.1　融合误差对比分析

在 $f = 100\mathrm{Hz}$、故障节点比例为 5%条件下，分别采用 ADAEWF 算法、AFWDF 算法和算术平均值法对连续 80 个周期的节点数据进行数据融合，融合曲线如图 6 所示。仿真结果表明：AFWDF 和算术平均值法的融合结果波动较大，而 ADAEWF 融合结果较平稳，在融合精度和可靠性方面 ADAEWF 算法有明显优势。其原因在于 ADAEWF 算法的异常数据预处理机制，排除了农田温度无线监测系统中传感器节点产生的瞬时性、故障性异常数据参与数据融合，有效提高了融合结果的可靠性；另一方面，通过节点数据与初次分批估计融合结果的相对方差对融合权值调整，并进行自适应估计加权数据融合，进一步降低了传感器测量误差对融合结果的影响，提高了融合精度和可靠性。

在上述仿真条件下，三种算法融合结果的均方误差随故障节点比例变化的曲线如图 7 所示。由图 7 可知,故障节点比例低于 10%时，三种算法均方误差均小于 0.7℃；故障节点比例高于 10%时，AFWDF 和算术平均值法的均方误差逐渐增大，当接近 40%时，两种算法的均方误差分别增至 2.3℃和 3.4℃；在故障

节点比例由 0%增大到 40%的过程中，ADAEWF 算法的均方误差始终小于 0.8℃。对比分析结果表明：在农田温度无线监测系统中，网络出现故障节点情况下，ADAEWF 算法融合结果误差较小，具有较高的可靠性。

3.2　网络生存期对比分析

在 $f = 20\text{Hz}$、故障节点比例为 10%条件下，网络存活节点数随算法运行轮数变化曲线如图 8 所示。由图 8 可知：AFWDF 和算术平均值法分别在 850 轮和 760 轮时出现首个死亡节点，而 ADAEWF 算法在 1300 轮时出现首个死亡节点；70%节点死亡时，AFWDF 和算术平均值法分别运行 1230 轮和 1110 轮，而 ADAEWF 算法运行 1670 轮。在农田温度无线监测系统中，70%节点死亡后系统无法有效完成监测任务，认为网络进入死亡期。综上所述，ADAEWF 算法的网络有效生存期分别是算术平均值法的 150%、AFWDF 算法的 135%，ADAEWF 算法能量使用效率更高。原因是 AFWDF 和算术平均值法均为周期性进行数据采集、发送、融合，能量消耗较大；而 ADAEWF 算法仅在用户查询数据，或者在检测到农田异常温度数据时，才进行数据采集和发送，其他时间节点处于低功耗的抑制状态；簇首在受到足够强度的异常数据刺激时才触发融合算法。

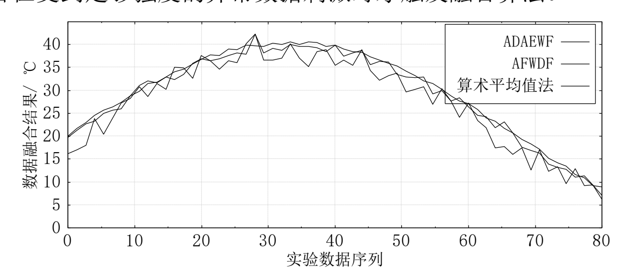


图 6　三种算法融合结果比较图

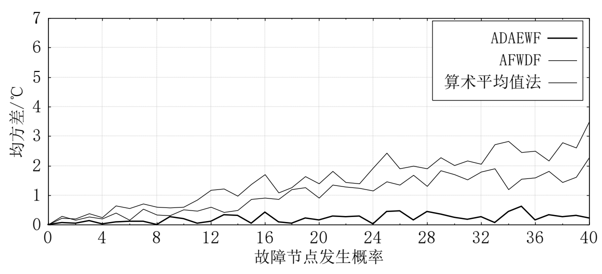


图 7　三种算法均方误差比较

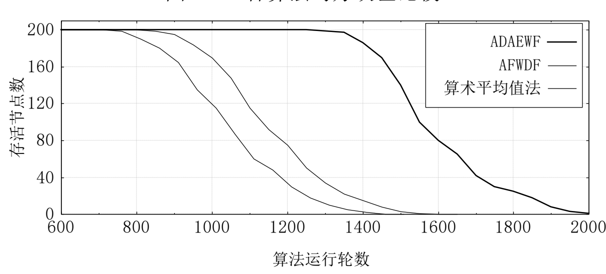


图 8　三种算法的存活节点数与轮数关系

4　结束语

本文对无线传感器网络数据融合进行研究，提出了异常数据预处理机制和自适应估计加权数据融合算法。不同于已有文献中连续或周期性触发的融合算法，ADAEWF 算法仅在系统检测到足够数量有效异常数据时才被触发运行，可以有效降低算法运行频率。此外，异常数据预处理机制排除了瞬时性、故障性异常数据参与数据融合，有效提高了融合精度。仿真结果表明：在农田温度无线监测系统仿真模型下，相较于算术平均值法和 AFWDF 算法，ADAEWF 算法具有较低的均方误差和较长的有效生存期。

参考文献：

[1] 李怀俊，张学习. 基于二维信息熵的无线传感网络簇内数据融合方法的研究 [J]. 计算机应用研究，2014, 31 (7): 2171-2174. (Li Huaijun, Zhang Xuexi, Research on data fusion method in cluster for wireless sensor network based on 2D entropy theory [J]. Application Research of Computers, 2014, 31 (7): 2171-2174.)

[2] 余修武，范飞生，周利兴，等. 无线传感器网络自适应预测加权数据融合算法 [J]. 传感技术学报，2017, 30 (5): 772-776. (Yu Xiuwu, Fan Feisheng, Zhou Lixing, *et al.* Adaptive Forecast Weighting Data Fusion Algorithm for Wireless Sensor Network [J]. Chinese Journal of Sensors and Actuators, 2017, 30 (5): 772-776.)

[3] 何友，王国宏，彭应宁，等. 多传感器信息融合及应用 [M]. 北京：电子工业出版社, 2000.

[4] 璩晶磊，李少波，张成龙. 基于模糊证据理论的多传感器数据融合算法 [J]. 仪表技术与传感器, 2017 (10): 118-122. (Qu Jinglei, Li Shaobo, Zhang Chenglong, Multisensor Data Fusion Algorithm Based on Fuzzy Evidence Theory [J]. Instrument Technique and Sensor, 2017 (10): 118-122.)

[5] Chhabra S, Singh D. Data fusion and data aggregation//summarization techniques in wsns: a review [J]. International Journal of Computer Applications, 2015, 121 (19): 21-30.

[6] Awang A, Agarwal S. Data aggregation using dynamic selection of aggregation points based on rssi for wireless sensor networks [J]. Wireless Personal Communications, 2015, 80 (2): 611-633.

[7] Rout R R, Ghosh S K. Adaptive data aggregation and energy efficiency using network coding in a clustered wireless sensor network: An analytical approach [M]. Elsevier Science Publishers B. V. 2014, 40: 65-75.

[8] Yeganeh M H, Yousefi H, Alinaghipour N, *et al.* RDAG: a structure-free real-time data aggregation protocol for wireless sensor networks [C]// Proc of the 17th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications. 2011, 1: 51-60.

[9] 侯鑫，张东文，钟鸣. 基于事件驱动和神经网络的无线传感器网络 数据融合算法研究 [J]. 传感技术学报，2014. 27 (1): 142-148. (Hou Xin, Zhang Dongwen, Zhong Ming, Data aggregation of wireless sensor network based on event-driven and neural network [J]. Chinese Journal of Sensors and Actuators, 2014. 27 (1): 142-148

[10] 邱立达，刘天键，傅平. 基于稀疏滤波的无线传感器网络数据融合 [J]. 电子测量与仪器学报, 2015, 29 (3): 352-357. (Qiu Lida, Liu Tianjian, Fu Ping, Data fusion in wireless sensor network based on sparse filtering [J]. Journal of Electronic Measurement and Instrumentation. 2015, 29 (3): 352-357.

[11] 谭德坤, 付雪峰, 赵嘉, 等. 基于异常数据驱动的 WSN 簇内数据融合方法 [J]. 传感技术学报, 2017, 30 (2): 306-312. (Tan Dekun, Fu Xuefeng, Zhao Jia, *et al.* A data aggregation method based on abnormal data-driven in clusters of wireless sensor networks [J]. Chinese Journal of Sensors and Actuators, 2017, 30 (2): 306-312

[12] 吕涛, 朱清新, 朱玉玉. 一种能耗均衡的无线传感器网络分簇算法 [J]. 计算机应用, 2012, 32 (11): 3107-3111. (Lyu Tao, Zhu Qingxin, Zhu Yuyu, Energy-balanced adaptive clustering algorithm for wireless sensor network [J]. Journal of Computer Applications. 2012, 32 (11): 3107-3111.

[13] 万叶晶, 叶继华, 江爱文. 一种基于时空相关性和异常检测的改进 WSN 节能策略 [J]. 传感技术学报, 2017, 30 (8): 1267-1273. (Wan Yejing, Ye Jihua, Jiang Aiwen, An improved wsn energy saving strategy based on spatio-temporal correlation and anomaly detection [J]. Chinese Journal of Sensors and Actuators, 2017, 30 (8): 1267-1273.)

[14] 费欢, 李光辉. 基于 K-means 聚类的 WSN 异常数据检测算法 [J]. 计算机工程, 2015, 41 (7): 124-128. (Fei Huan, Li Guanghui, Abnormal Data Detection Algorithm for WSN Based on K-means Clustering [J]. Computer Engineering, 2015, 41 (7): 124-128.)

[15] Ding M, Chen D, Xing K, *et al.* Localized Fault-Tolerant Event Boundary Detection in Sensor Networks [C]// Proc of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, 2005, 2: 902-913.

[16] 刘思峰, 谢乃明, FORREST Jeffery. 基于相似性和接近性视角的新型灰色关联分析模型 [J]. 系统工程理论与实践, 2010, 30 (5): 881-887. (Liu Sifeng, Xie Naiming, FORREST Jeffery, On new models of grey incidence analysis based on visual angle of similarity and nearness [J]. Systems Engineering-Theory & Practice, 2010, 30 (5): 881-887.)

[17] 蒋婷, 滕召胜, 顾红艳, 等. 基于 EMD 与分批估计的动态称量快速融合方法 [J]. 仪器仪表学报, 2015, 36 (6): 1406-1414. (Jiang Ting, Teng Zhaosheng, Gu Hongyan, *et al.* Fast fusion method of dynamic weighing based on EMD and batch estimation [J]. Chinese Journal of Scientific Instrument. 2015, 36 (6): 1406-1414.